The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Parameter-Free Centralized Multi-Task Learning for Characterizing Developmental Sex Differences in Resting State Functional Connectivity

Xiaofeng Zhu, Hongming Li, Yong Fan[*†]
Department of Radiology
Perelman School of Medicine
University of Pennsylvania, Philadelphia, PA 19104, USA

Abstract

In contrast to most existing studies that typically characterize the developmental sex differences using analysis of variance or equivalently multiple linear regression, we present a parameter-free centralized multi-task learning method to identify sex specific and common resting state functional connectivity (RSFC) patterns underlying the brain development based on resting state functional MRI (rs-fMRI) data. Specifically, we design a novel multi-task learning model to characterize sex specific and common RSFC patterns in an age prediction framework by regarding the age prediction for males and females as separate tasks. Moreover, the importance of each task and the balance of these two patterns, respectively, are automatically learned in order to make the multi-task learning robust as well as free of tunable parameters, *i.e.,* parameter-free for short. Our experimental results on synthetic datasets verified the effectiveness of our method with respect to prediction performance, and experimental results on rs-fMRI scans of 1041 subjects (651 males) of the Philadelphia Neurodevelopmental Cohort (PNC) showed that our method could improve the age prediction on average by 5.82% with statistical significance than the best alternative methods under comparison, in addition to characterizing the developmental sex differences in RSFC patterns.

Introduction

Adolescence is the developmental period during which functional brain maturation interacts with sexual divergence in social, behavior, and biological changes (Kochhann et al. 2017). Particularly, sex differences are prominent in behavior and have been studying for a long time (Hardee et al. 2017; Qian et al. 2015a). For instance, females are superior at social cognition and recognition memory than males, while males perform better in visuospatial and motor tasks than females (Gur and Gur 2016).

It is of great importance to understand the neural origins of the developmental sex differences in behavior as both sex differences and brain structure/function are shaped during adolescence to support the brain development and neuropsychiatric disorders typically begin in adolescence and are linked to aberrations in neurodevelopment (Di Martino et al. 2014). Recently, neuroimaging measures, including both structural and functional neuroimaging measures, have been adopted as surrogate neural variables for exploring neural origins of the developmental sex differences (Ingalhalikar et al. 2014; Alarcón et al. 2015). Particularly, magnetic resonance imaging (MRI) is a widely used technique for sex differences analysis (Gur and Gur 2016). Several studies on sex differences via analyzing gray matter (GM) and white matter (WM) of the brain MRI data have reported that females have smaller brain volume and cerebral spinal fluid (CSF) volume than males, and CSF volume changes faster in males than in females (Blakemore, Burnett, and Dahl 2010).

Resting state functional MRI (rs-fMRI) provides task-independent and relatively reproducible biomarkers of functional coherence of activity in different brain regions (Fox and Raichle 2007). Using resting state functional connectivity (RSFC) analytic techniques, we are able to investigate the brain functional organization of both typical brain development and neuropsychiatric disorders (Fox and Raichle 2007; Di Martino et al. 2014) and we can also characterize the brain state at an individual subject level based on the RSFC measures using pattern recognition techniques (Fan and Davatzikos 2017). Recent studies have demonstrated that RSFC measures are more accurate than cognitive profiles for both sex classification and sex differences characterization, and males exhibit weaker nucleus accumbens functional connectivity than females in adolescent brains (Müller-Oehring et al. 2017).

Since brain development differs between males and females across adolescence (Ingalhalikar et al. 2014; Alarcón et al. 2015), the interaction of age and sex on RSFC has been investigated using analysis of variance (ANOVA) or equivalently multiple linear regression for exploring neural origins of sex differences in the developing brain (Alarcón et al. 2015). However, such analytic tools are typically adopted in a univariate statistical analysis, not equipped to characterize multivariable relationships, such as the interaction of age and sex on the functional brain network which is typically characterized by a set of edgewise RSFC measures (Alarcón et al. 2015).

In order to robustly characterize the interaction of sex and age on the RSFC measures in a multivariate analy-

[*]Corresponding author: yong.fan@uphs.upenn.edu.

[†]This work was supported in part by National Institutes of Health grants (Nos. EB022573, CA189523, MH107703, DA039215, and DA039002).

sis setting, we present a parameter-free centralized multi-task learning method to identify sex specific and common RSFC patterns underlying the brain development based on rs-fMRI data in an age prediction framework. Specifically, our method explicitly identifies two sets of RSFC measures, namely male specific and female specific RSFC measures, which contribute differently to the age prediction by modeling the age prediction for males and females as separate tasks in a multi-task learning framework. At the same time, the male specific and female specific RSFC measures are regularized to be sparse and centralized to have a shared, sex common pattern by optimizing a square root objective function that models difference between measured and predicted ages with an $\ell_{2,1}$-norm regularization (Hu et al. 2017; Zhu et al. 2017b; 2016b; Peng and Fan 2016). The importance of each task and the balance of these two patterns, respectively, are automatically learned to make the multi-task learning free of tunable parameters, *i.e.,* parameter-free for short. As a result, the parameter-free, centralized, sparse multi-task learning makes our method fast and robust to noisy measures/outliers. Akin to the ANOVA method, the sex specific RSFC patterns characterize sex differences while the sex common RSFC patterns characterize the interaction effect of sex and age on the developing RSFC measures. Furthermore, our method may make these two patterns collaboratively help each other to improve the prediction performance of each task and discover more interesting patterns which cannot be found in a model built on data from only one sex group.

We have validated our method based on synthetic and real datasets. Our experimental results on synthetic datasets verified the effectiveness of our method with respect to the age prediction performance. We also investigated the developmental sex differences in RSFC measures based on rs-fMRI data of 1041 subjects (651 males) of the Philadelphia Neurodevelopmental Cohort (PNC) dataset (Li, Satterthwaite, and Fan 2017). Our experimental results showed that our method could improve the age prediction on average by 5.82% with statistical significance than the best alternative methods under comparison, including state-of-the-art multi-task learning methods, sex specific age prediction models, and an age prediction model with sex as a feature, in addition to characterizing the developmental sex differences in RSFC patterns.

Methods

In this paper, we denote matrices as boldface uppercase letters, vectors as boldface lowercase letters, and scalars as normal italic letters. For a matrix $\mathbf{X} = [x_{ij}]$, its *i*-th row and *j*-th column are denoted as $\mathbf{x}^i$ and $\mathbf{x}_j$, respectively. We also denote the Frobenius norm and the ℓ_1-norm of a matrix $\mathbf{X}$ as $\|\mathbf{X}\|_F = \sqrt{\sum_j \|\mathbf{x}_j\|_2^2}$ and $\|\mathbf{X}\|_1 = \sum_{ij} |x_{ij}|$, respectively. We denote the transpose operator, the trace operator, and the inverse of a matrix $\mathbf{X}$ as $\mathbf{X}^T$, $tr(\mathbf{X})$, and $\mathbf{X}^{-1}$, respectively.

Sparse feature selection

Given an RSFC feature matrix $\mathbf{X} \in \mathbb{R}^{n\times d}$ and its associated age vector $\mathbf{y} \in \mathbb{R}^n$, where n and d, respectively, are the number of subjects and the dimensionality of RSFC features, we assume that there is a linear relationship between the RSFC features $\mathbf{X}$ and the age vector $\mathbf{y}$. We then use the least square loss function to measure their similarity or relationship via following formulation:

$$\min_{\mathbf{w}} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2, \tag{1}$$

where the coefficient vector $\mathbf{w} \in \mathbb{R}^d$ maps $\mathbf{X}$ to $\mathbf{y}$ for achieving the minimal prediction residual $\|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2$, and $\mathbf{X}\mathbf{w}$ is the prediction of $\mathbf{y}$.

The least square regression in Eq. (1) has a closed form solution, *i.e.,* $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. However, the inverse operator is often ill-posed when dealing with high-dimensional data (Zhu et al. 2017a; Peng and Fan 2017a), *e.g.,* the large number of d ($d > n$) in the present study. In this case, a regularization is always recommended (Peng and Fan 2017b; Qian et al. 2015b; Chang et al. 2014; Zhu et al. 2017c). On the other hand, not all the features (*i.e.,* functional connectivity measures between different nodes of the brain) are predictive for the brain development, *i.e.,* the age prediction. To address above issues, we employ a sparse regularization to solve the ill-posed issue and select important features. We thus have following formulation:

$$\min_{\mathbf{w}} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_1, \tag{2}$$

where $\lambda \geq 0$ is a nonnegative tuning parameter, and a large value of λ encourages sparsity of the model.

After solving Eq. (2), the features with zero coefficients in $\mathbf{w}$ are regarded as unimportant features while the remaining features with nonzero coefficients are regarded as important features (Yang et al. 2015; Zhang et al. 2017). In this way, we may use Eq. (2) to conduct feature selection in the individual groups (*i.e.,* the female group and the male group, respectively). We can also concatenate the data points in different groups to form a large dataset to build an age prediction model for each group.

Parameter-free centralized multi-task learning

According to existing studies of the developmental sex difference in RSFC patterns (Alarcón et al. 2015) and our experimental results summarized in Table 4, we observe that males and females develop differently with respect to their RSFC patterns, but share common developmental patterns. Therefore, we propose to simultaneously identify sex specific RSFC patterns and sex common patterns in a uniformed framework, with the expectation that these sex groups collaboratively help each other to improve the prediction performance and discover more interesting patterns, which cannot be found in a model built with only one sex group.

To do this, we regard the prediction of each group (*i.e.,* the female group and the male group, respectively) as one task. We then denote the RSFC feature matrices as $\mathbf{X}_1 \in \mathbb{R}^{n_1\times d}$ and $\mathbf{X}_2 \in \mathbb{R}^{n_2\times d}$ (where n_1 and n_2, respectively, are the number of the subjects of these two tasks), and their corresponding age vectors as $\mathbf{y}_1 \in \mathbb{R}^{n_1}$ and $\mathbf{y}_2 \in \mathbb{R}^{n_2}$, respectively. It is noteworthy that our method can deal with cases where both the number of the subjects and the dimensions of the features are different in these two tasks. We further use

the least square loss function to achieve the minimal prediction error of all the tasks and to select the informative features with following multi-task learning formulation:

$$\min_{\mathbf{w}_t} \sum_{t=1}^{k} \|\mathbf{y}_t - \mathbf{X}_t\mathbf{w}_t\|_2^2 + \gamma\|\mathbf{W}\|_1, \tag{3}$$

where $\mathbf{W} = [\mathbf{w}_1, ..., \mathbf{w}_t] \in \mathbb{R}^{d\times k}$ and γ is a nonnegative tuning parameter to control the sparsity ratio of $\mathbf{W}$.

In this study, to achieve the goal that two different tasks collaboratively help each other, we use a centralized regularization to penalize the variance of the coefficient vectors (*i.e.,* $\mathbf{w}_t$, $t = 1, ..., k$) by optimizing following objective function:

$$\begin{aligned}\min_{\mathbf{w}_t,\bar{\mathbf{w}}} \sum_{t=1}^{k} \sqrt{\|\mathbf{y}_t - \mathbf{X}_t\mathbf{w}_t\|_2^2 + \|\mathbf{w}_t - \bar{\mathbf{w}}\|_{2,1}} \\ + \gamma\|\mathbf{W}\|_1,\end{aligned} \tag{4}$$

To solve the optimization problem of Eq. (4), *i.e.,* optimizing the variables $\mathbf{w}_t$ ($t = 1, ..., k$) and $\bar{\mathbf{w}}$, we compute the derivatives of the square root in Eq. (4) and obtain the following formulation

$$\begin{cases} \min_{\mathbf{w}_t,\bar{\mathbf{w}}} \sum_{t=1}^{k} \alpha_t(\|\mathbf{y}_t - \mathbf{X}_t\mathbf{w}_t\|_2^2 + \|\mathbf{w}_t - \bar{\mathbf{w}}\|_{2,1}) \\ \quad + \gamma\|\mathbf{W}\|_1, & \text{(5a)} \\ \alpha_k = \dfrac{1}{2\sqrt{\|\mathbf{y}_t - \mathbf{X}_t\mathbf{w}_t\|_2^2 + \|\mathbf{w}_t - \bar{\mathbf{w}}\|_{2,1}}}. & \text{(5b)} \end{cases}$$

Algorithm 1: The optimization algorithm for Eq. (4).

Input: γ, $\mathbf{X}_t$, and $\mathbf{y}_t$ ($t = 1, ..., k$);
Output: α_t, $\bar{\mathbf{w}}$, and $\mathbf{w}_t$ ($t = 1, ..., k$);

```
Initialize α_t as α_t = 1/k (t = 1, ..., k);
repeat
    Optimize Eq. (5a);
    Update α_t via Eq. (5b);
until Eq. (4) converges;
```

Given α_k in Eq. (5b), we can optimize Eq. (5a) via sequentially taking the derivative with respect to the variables $\mathbf{w}_t$ ($t = 1, ..., k$), and $\bar{\mathbf{w}}$. However, the value of α_k is dependent on the optimization of these variables, which are also dependent on the value of α_k. In this study, we propose to solve the original objective function in Eq. (4) via alternatively updating Eq. (5b) and Eq. (5a), until Algorithm 1 converges.

In Algorithm 1, the optimization of Eq. (5b) is straightforward, and we therefore focus on the optimization of Eq. (5a) via alternatively solving following two formulations iteratively until the objective function value of Eq. (5a) is stable

$$\begin{cases} \min_{\mathbf{w}_t,\bar{\mathbf{w}}} \sum_{t=1}^{k} \alpha_t(\|\mathbf{y}_t - \mathbf{X}_t\mathbf{w}_t\|_2^2 + \beta_t\|\mathbf{w}_t - \bar{\mathbf{w}}\|_2^2) \\ \quad + \gamma\|\mathbf{W}\|_1, & \text{(6a)} \\ \beta_k = \dfrac{1}{2\sqrt{\|\mathbf{w}_t - \bar{\mathbf{w}}\|_2}} & \text{(6b)} \end{cases}$$

We arrange Eq. (6a) to obtain the following formulation:

$$\min_{\mathbf{w}_t,\bar{\mathbf{w}}} \sum_{t=1}^{k} \|\hat{\mathbf{y}}_t - \hat{\mathbf{X}}_t\mathbf{w}_t\|_2^2 + \gamma\|\mathbf{W}\|_1 \tag{7}$$

where $\hat{\mathbf{y}}_t = [\sqrt{\alpha_k}\mathbf{y}_t^T, \sqrt{\alpha_k\beta_k}\bar{\mathbf{w}}^T]^T \in \mathbb{R}^{(n+d)\times 1}$, $\hat{\mathbf{X}}_t = [\sqrt{\alpha_k}\mathbf{X}_t^T, \sqrt{\alpha_k\beta_k}\mathbf{I}]^T \in \mathbb{R}^{(n+d)\times d}$, and $\mathbf{I} \in \mathbb{R}^{d\times d}$ is an identity matrix. Eq. (7) is a standard objective function of sparse multi-task learning, which can be solved by the toolbox MALSAR (Zhou, Chen, and Ye 2011).

The optimization of our objective function Eq. (4) may obtain a local optimal solution. However, both the prediction and the convergence of our method were insensitive to the initialization because our method achieved good prediction performance and fast convergence by only initializing $\mathbf{w}_t$ with the least square results, *i.e.,* $(\mathbf{X}_t^T\mathbf{X}_t + 0.001 * \mathbf{I})^{-1}\mathbf{X}_t^T\mathbf{y}_t$, and setting $\bar{\mathbf{w}}$ to the average of all $\mathbf{w}_t$, $\alpha_t = \frac{1}{k}$, and $\beta_t = \frac{1}{k}$, $t = 1, ..., k$.

Our objective function in Eq. (4) brings several advantages. Firstly, once Eq. (6a) achieves convergence, the values of α_k in Eq. (5b) and β_t in Eq. (6b), respectively, can be regarded as weights of the tasks and the centralized regularization. It is noteworthy that both α_k and β_t are automatically obtained without tuning parameters, *i.e.,* parameter-free. Moreover, if the t-task is important, then the prediction error of both the loss function and the centralized regularization are small. This indicates that our parameter-free method is meaningful. Secondly, $\bar{\mathbf{w}}$ is the mean vector of $\mathbf{w}_t$ ($t = 1, ..., k$), while $\alpha_t\beta_t$ is the weight to reduce the variance of $\mathbf{w}_t$ in the term $\alpha_t\beta_t\|\mathbf{w}_t - \bar{\mathbf{w}}\|_2^2$ in Eq. (6b), *i.e.,* making all the tasks similar. Specifically, $\alpha_t\beta_t$ is used to measure the diversity and the flexibility of $\mathbf{X}_t$. If $\mathbf{X}_t$ is more similar to other tasks, the value of $\alpha_t\beta_t$ should be bigger to push $\mathbf{w}_t$ closer to $\bar{\mathbf{w}}$. In this case, $\mathbf{X}_t$ has less flexibility (*i.e.,* sex specific patterns) and more stability (*i.e.,* sex common patterns). Moreover, the parameter-free optimization automatically balances contributions of sex specific patterns and sex common patterns to the age prediction.

Convergence Analysis

The convergence of Algorithm 1 and Eq. (6a) is provided as following. Particularly, the convergence of Eq. (6a) has been proved in (Zhou, Chen, and Ye 2011). Algorithm 1 is a special case of the Iteratively Re-weighted Least Square (IRLS) framework (Daubechies et al. 2010; Zhu et al. 2017a), and its convergence and effectiveness have been theoretically verified. To do this, we have following lemma (Hu et al. 2017; Zhu et al. 2017b).

Lemma 1. *For any positive real numbers u and v, following inequality always holds:*

$$\sqrt{u} - \frac{u}{2\sqrt{v}} \leq \sqrt{v} - \frac{v}{2\sqrt{v}}. \tag{8}$$

Theorem 1. *The objective function value in Eq. (4) monotonically decreases until Algorithm 1 converges.*

Proof. We denote $\hat{\mathbf{w}}_t$, $\hat{\bar{\mathbf{w}}}$, and $\hat{\alpha}_t$ as the updated $\mathbf{w}_t$, $\bar{\mathbf{w}}$, and α_t in each iteration, and then present the convergence analysis of Algorithm 1 via following three steps.

- Obtain $\hat{\mathbf{w}}_t$ ($t = 1, ..., k$) while fixing $\bar{\mathbf{w}}$, and α_t.
We change Eq. (4) with respect to $\mathbf{w}_t$ ($t = 1, ..., k$) to:

$$\hat{\mathbf{w}}_t = \min_{\mathbf{w}_t} \sqrt{\|\mathbf{y}_t - \mathbf{X}_t\mathbf{w}_t\|_2^2 + \|\mathbf{w}_t - \bar{\mathbf{w}}\|_{2,1}} + \gamma\|[\mathbf{w}_1, ..., \mathbf{w}_t, ..., \mathbf{w}_k]\|_1 \tag{9}$$

According to Eq. (5b), Eq. (6a), and (Zhou, Chen, and Ye 2011), for each *i* ($i = 1, ..., d$), we have

$$\begin{aligned}&\frac{\|\mathbf{y}_t-(\mathbf{X}_t)_{.,i}(\hat{\mathbf{w}}_t)_i\|_2^2+\|(\hat{\mathbf{w}}_t)_i-(\bar{\mathbf{w}})_i\|_{2,1}}{2\sqrt{\|\mathbf{y}_t-(\mathbf{X}_t)_{.,i}(\mathbf{w}_t)_i\|_2^2+\|(\mathbf{w}_t)_i-(\bar{\mathbf{w}})_i\|_{2,1}}} + \gamma\frac{((\hat{\mathbf{w}}_t)_i)^2}{2\|(\mathbf{W})_{.,i}\|} \\ \leq &\frac{\|\mathbf{y}_t-(\mathbf{X}_t)_{.,i}(\mathbf{w}_t)_i\|_2^2+\|(\mathbf{w}_t)_i-(\bar{\mathbf{w}})_i\|_{2,1}}{2\sqrt{\|\mathbf{y}_t-(\mathbf{X}_t)_{.,i}(\mathbf{w}_t)_i\|_2^2+\|(\mathbf{w}_t)_i-(\bar{\mathbf{w}})_i\|_{2,1}}} + \gamma\frac{((\mathbf{w}_t)_i)^2}{2\|(\mathbf{W})_{.,i}\|}\end{aligned} \tag{10}$$

where $(\mathbf{X}_t)_{.,i}$ is the *i*-column of the matrix $\mathbf{X}_t$ and $(\hat{\mathbf{w}}_t)_i$ is the *i*-th element of the vector $\hat{\mathbf{w}}_t$. According to Lemma 1, we have

$$\begin{aligned}&\sqrt{\mathbf{y}_t - (\mathbf{X}_t)_{.,i}(\hat{\mathbf{w}}_t)_i\|_2^2 + \|(\hat{\mathbf{w}}_t)_i - (\bar{\mathbf{w}})_i\|_{2,1}} + \gamma(\hat{\mathbf{w}}_t)_i \\ &-(\frac{\|\mathbf{y}_t-(\mathbf{X}_t)_{.,i}(\hat{\mathbf{w}}_t)_i\|_2^2+\|(\hat{\mathbf{w}}_t)_i-(\bar{\mathbf{w}})_i\|_{2,1}}{2\sqrt{\|\mathbf{y}_t-(\mathbf{X}_t)_{.,i}(\mathbf{w}_t)_i\|_2^2+\|(\mathbf{w}_t)_i-(\bar{\mathbf{w}})_i\|_{2,1}}} + \gamma\frac{((\hat{\mathbf{w}}_t)_i)^2}{2\|(\mathbf{W})_{.,i}\|}) \\ \leq &\sqrt{\|\mathbf{y}_t - (\mathbf{X}_t)_{.,i}(\mathbf{w}_t)_i\|_2^2 + \|(\mathbf{w}_t)_i - (\bar{\mathbf{w}})_i\|_{2,1}} + \gamma(\mathbf{w}_t)_i \\ &-(\frac{\|\mathbf{y}_t-(\mathbf{X}_t)_{.,i}(\mathbf{w}_t)_i\|_2^2+\|(\mathbf{w}_t)_i-(\bar{\mathbf{w}})_i\|_{2,1}}{2\sqrt{\|\mathbf{y}_t-(\mathbf{X}_t)_{.,i}(\mathbf{w}_t)_i\|_2^2+\|(\mathbf{w}_t)_i-(\bar{\mathbf{w}})_i\|_{2,1}}} + \gamma\frac{((\mathbf{w}_t)_i)^2}{2\|(\mathbf{W})_{.,i}\|})\end{aligned} \tag{11}$$

By combining Eq. (10) with Eq. (11), we obtain

$$\begin{aligned}&\sqrt{\mathbf{y}_t - (\mathbf{X}_t)_{.,i}(\hat{\mathbf{w}}_t)_i\|_2^2 + \|(\hat{\mathbf{w}}_t)_i - (\bar{\mathbf{w}})_i\|_{2,1}} + \gamma(\hat{\mathbf{w}}_t)_i \\ \leq &\sqrt{\|\mathbf{y}_t - (\mathbf{X}_t)_{.,i}(\mathbf{w}_t)_i\|_2^2 + \|(\mathbf{w}_t)_i - (\bar{\mathbf{w}})_i\|_{2,1}} + \gamma(\mathbf{w}_t)_i\end{aligned} \tag{12}$$

After summing all the *i*, we obtain

$$\begin{aligned}&\sqrt{\|\mathbf{y}_t - \mathbf{X}_t\hat{\mathbf{w}}_t\|_2^2 + \|\hat{\mathbf{w}}_t - \bar{\mathbf{w}}\|_{2,1}} + \gamma\|\hat{\mathbf{w}}_t\|_1 \\ \leq &\sqrt{\|\mathbf{y}_t - \mathbf{X}_t\mathbf{w}_t\|_2^2 + \|\mathbf{w}_t - \bar{\mathbf{w}}\|_{2,1}} + \gamma\|\hat{\mathbf{w}}_t\|_1\end{aligned} \tag{13}$$

After summing all the tasks, we have

$$\begin{aligned}&\sum_{t=1}^{k}\sqrt{\|\mathbf{y}_t - \mathbf{X}_t\hat{\mathbf{w}}_t\|_2^2 + \|\hat{\mathbf{w}}_t - \bar{\mathbf{w}}\|_{2,1}} + \gamma\|\hat{\mathbf{W}}\|_1 \\ \leq &\sum_{t=1}^{k}\sqrt{\|\mathbf{y}_t - \mathbf{X}_t\mathbf{w}_t\|_2^2 + \|\mathbf{w}_t - \bar{\mathbf{w}}\|_{2,1}} + \gamma\|\mathbf{W}\|_1\end{aligned} \tag{14}$$

- Obtain $\hat{\bar{\mathbf{w}}}$ while fixing $\hat{\mathbf{w}}_t$ ($t = 1, ..., k$), and α_t.
We change Eq. (4) with respect to $\bar{\mathbf{w}}$ to:

$$\hat{\bar{\mathbf{w}}} = \min_{\bar{\mathbf{w}}} \sqrt{\sum_{t=1}^{k}\|\mathbf{y}_t - \mathbf{X}_t\hat{\mathbf{w}}_t\|_2^2 + \|\hat{\mathbf{w}}_t - \bar{\mathbf{w}}\|_{2,1}} + \gamma\|\hat{\mathbf{W}}\|_1 \tag{15}$$

According to Eq. (5b) and Eq. (15), we have

$$\begin{aligned}&\frac{\sum_{t=1}^{k}(\|\mathbf{y}_t-\mathbf{X}_t\hat{\mathbf{w}}_t\|_2^2+\|\hat{\mathbf{w}}_t-\hat{\bar{\mathbf{w}}}\|_{2,1})}{2\sqrt{\sum_{t=1}^{k}(\|\mathbf{y}_t-\mathbf{X}_t\hat{\mathbf{w}}_t\|_2^2+\|\hat{\mathbf{w}}_t-\hat{\bar{\mathbf{w}}}\|_{2,1})}} + \gamma\|\hat{\mathbf{W}}\|_1 \\ \leq &\frac{\sum_{t=1}^{k}(\|\mathbf{y}_t-\mathbf{X}_t\hat{\mathbf{w}}_t\|_2^2+\|\hat{\mathbf{w}}_t-\bar{\mathbf{w}}\|_{2,1})}{2\sqrt{\sum_{t=1}^{k}(\|\mathbf{y}_t-\mathbf{X}_t\hat{\mathbf{w}}_t\|_2^2+\|\hat{\mathbf{w}}_t-\bar{\mathbf{w}}\|_{2,1})}} + \gamma\|\hat{\mathbf{W}}\|_1\end{aligned} \tag{16}$$

According to Lemma 1, we have

$$\begin{aligned}&\sqrt{\sum_{t=1}^{k}\|\mathbf{y}_t - \mathbf{X}_t\hat{\mathbf{w}}_t\|_2^2 + \|\hat{\mathbf{w}}_t - \hat{\bar{\mathbf{w}}}\|_{2,1}} - \\ &\frac{\sum_{t=1}^{k}\|\mathbf{y}_t-\mathbf{X}_t\hat{\mathbf{w}}_t\|_2^2+\|\hat{\mathbf{w}}_t-\hat{\bar{\mathbf{w}}}\|_{2,1}}{2\sqrt{\sum_{t=1}^{k}(\|\mathbf{y}_t-\mathbf{X}_t\hat{\mathbf{w}}_t\|_2^2+\|\hat{\mathbf{w}}_t-\bar{\mathbf{w}}\|_{2,1})}} + \gamma\|\hat{\mathbf{W}}\|_1 \\ \leq &\sqrt{\sum_{t=1}^{k}\|\mathbf{y}_t - \mathbf{X}_t\hat{\mathbf{w}}_t\|_2^2 + \|\hat{\mathbf{w}}_t - \bar{\mathbf{w}}\|_{2,1}} - \\ &\frac{\sum_{t=1}^{k}\|\mathbf{y}_t-\mathbf{X}_t\hat{\mathbf{w}}_t\|_2^2+\|\hat{\mathbf{w}}_t-\bar{\mathbf{w}}\|_{2,1}}{2\sqrt{\sum_{t=1}^{k}(\|\mathbf{y}_t-\mathbf{X}_t\hat{\mathbf{w}}_t\|_2^2+\|\hat{\mathbf{w}}_t-\bar{\mathbf{w}}\|_{2,1})}} + \gamma\|\hat{\mathbf{W}}\|_1\end{aligned} \tag{17}$$

By combining Eq. (16) with Eq. (17), we obtain

$$\begin{aligned}&\sum_{t=1}^{k}\sqrt{\|\mathbf{y}_t - \mathbf{X}_t\hat{\mathbf{w}}_t\|_2^2 + \|\hat{\mathbf{w}}_t - \hat{\bar{\mathbf{w}}}\|_{2,1}} + \gamma\|\hat{\mathbf{W}}\|_1 \\ \leq &\sum_{t=1}^{k}\sqrt{\|\mathbf{y}_t - \mathbf{X}_t\hat{\mathbf{w}}_t\|_2^2 + \|\hat{\mathbf{w}}_t - \bar{\mathbf{w}}\|_{2,1}} + \gamma\|\hat{\mathbf{W}}\|_1\end{aligned} \tag{18}$$

By combining Eq. (14) with Eq. (18), we have

$$\begin{aligned}&\sum_{t=1}^{k}\sqrt{\|\mathbf{y}_t - \mathbf{X}_t\hat{\mathbf{w}}_t\|_2^2 + \|\hat{\mathbf{w}}_t - \hat{\bar{\mathbf{w}}}\|} + \gamma\|\hat{\mathbf{W}}\|_1 \\ \leq &\sum_{t=1}^{k}\sqrt{\|\mathbf{y}_t - \mathbf{X}_t\mathbf{w}_t\|_2^2 + \|\mathbf{w}_t - \bar{\mathbf{w}}\|_{2,1}} + \gamma\|\mathbf{W}\|_1\end{aligned} \tag{19}$$

- Obtain $\hat{\alpha}_t$ while fixing $\hat{\bar{\mathbf{w}}}$ and $\hat{\mathbf{w}}_t$ ($t = 1, ..., k$).
According to Eq. (5b), $\hat{\alpha}_t$ has a closed form solution, and thus Eq. (19) will be held.

Finally, the iterative optimization in Algorithm 1 will monotonically decrease the objective function value of Eq. (4) in each iteration until Algorithm 1 converges to a local optimization solution of Eq. (4). □

Results

Experimental Settings

To evaluate our method, we compared it with following methods based on data from distinct groups, *e.g.,* males and females, in a regression prediction setting.

Firstly, we conducted least square linear regression (Regression for short) (Friedman, Hastie, and Tibshirani 2001)

and Lasso (Tibshirani 1996) to build group specific prediction models upon the data points from each group separately.

Secondly, we built a single prediction model on pooled data points from all groups, regardless their group differences using either least square linear regression (ConRegression for short) or Lasso (ConLasso for short). And then, we estimated the performance of the prediction model for each group separately.

Thirdly, we used the group information as an extra feature to build a single prediction model using Lasso, referred to as MixLasso for short. We also estimated the performance of the prediction model for each group separately.

Lastly, by regarding the prediction for each group as a different task, we carried out multi-task learning with the standard multi-task learning method (MKL for short) (Zhu et al. 2016a; Argyriou, Evgeniou, and Pontil 2007; Zhu, Suk, and Shen 2014). Different from our method, the MKL method does not consider the weight of different tasks or the centralized regularization.

All the methods were evaluated using 10-fold cross-validation. We repeated the whole process 20 times to avoid possible bias during data partitioning for cross-validation. The final results were average of all 20 experiments. For the model selection, the parameters of the regression method and the Lasso were tuned based on cross-validation, and we carried out a grid searching on the parameter $\lambda \in \{10^{-3}, ..., 10^{3}\}$ for MKL and our method.

In all experiments, the prediction performance was measured using Correlation Coefficient (CC) between the predicted and real values.

Data

Synthetic data We generated four synthetic datasets by setting different values for the number of groups (k), the number of features (d), and the number of samples (n). Specifically, each element of $\mathbf{X}_t$ was sampled independent and identically distributed (i.i.d) from Gaussian distribution $N(0, 1)$ and then each feature was normalized to have a unit scale. Each element of the true $\mathbf{w}_t$ was sampled i.i.d from the uniform distribution in the interval $[-5, 5]$. We randomly set zeros to 20% rows of $\mathbf{W}$ and randomly replaced 60% elements of the remaining nonzero elements with zeros. Finally, noise $\boldsymbol{\delta}_t$ sampled i.i.d from Gaussian distribution $N(0, 10)$ was added to the features. The response was computed as $\mathbf{y}_t = \mathbf{X}_t\mathbf{w}_t + \boldsymbol{\delta}_t$.

PNC The Philadelphia Neurodevelopmental Cohort (PNC) is a collaboration between the Center for Applied Genomics at Childrens Hospital of Philadelphia (CHOP) and the Brain Behavior Laboratory at the University of Pennsylvania (Penn) (Satterthwaite et al. 2014). We followed (Li, Satterthwaite, and Fan 2017) to conduct data preprocessing. All data in this study were acquired on the same scanner (Siemens Tim Trio 3 Tesla, Erlangen, Germany; 32-channel head coil) using the same imaging sequences. One T1 scan was acquired prior to the rsfMRI scan with 124 time-points for each subject and the T1 images were processed using Freesurfer. Each rsfMRI scan was first registered to its corresponding T1 image, and then was projected to the fsaverage surface space via FreeSurfer after preprocessed using an optimized confound regression procedure (Li, Satterthwaite, and Fan 2017). Finally, we obtained preprocessed rsfMRI scans of 1401 subjects aged 8-22 years (651 males). We examined sex differences in 1540 edgewise RSFC measures within a brain network with 56 nodes described in (Li, Satterthwaite, and Fan 2017).

Table 1: CC results of Data 1 (k = 30, n = 1000, d = 400) and Data 2 (k = 50, n = 1000, d = 2000).

Methods	Data 1	Data 2
ConRegression	0.735 ± 0.025	0.676 ± 0.022
ConLasso	0.745 ± 0.027	0.698 ± 0.025
MixLasso	0.749 ± 0.031	0.704 ± 0.023
MKL	0.766 ± 0.032	0.711 ± 0.015
Proposed	0.799 ± 0.012	0.763 ± 0.015

Table 2: CC results of Data 3 (k = 2, n = 1000, d = 400)

Methods	Task 1	Task2
Regression	0.877 ± 0.173	0.851 ± 0.176
Lasso	0.889 ± 0.182	0.865 ± 0.195
ConRegression	0.878 ± 0.152	0.862 ± 0.232
ConLasso	0.890 ± 0.123	0.872 ± 0.214
MixLasso	0.909 ± 0.203	0.885 ± 0.167
MKL	0.923 ± 0.111	0.884 ± 0.202
Proposed	0.945 ± 0.112	0.921 ± 0.208

Table 3: CC results of Data 4 (k = 2, n = 1000, d = 2000)

Methods	Task 1	Task2
Regression	0.758 ± 0.054	0.801 ± 0.027
Lasso	0.767 ± 0.015	0.804 ± 0.031
ConRegression	0.785 ± 0.084	0.801 ± 0.047
ConLasso	0.788 ± 0.044	0.804 ± 0.028
MixLasso	0.796 ± 0.036	0.811 ± 0.035
MKL	0.805 ± 0.045	0.812 ± 0.045
Proposed	0.837 ± 0.024	0.842 ± 0.018

Table 4: CC results of PNC data

Methods	Males	Females
Regression	0.562 ± 0.017	0.397 ± 0.022
Lasso	0.566 ± 0.014	0.400 ± 0.025
ConRegression	0.566 ± 0.014	0.421 ± 0.024
ConLasso	0.567 ± 0.013	0.438 ± 0.021
MixLasso	0.571 ± 0.013	0.414 ± 0.017
MKL	0.581 ± 0.016	0.452 ± 0.028
Proposed	0.599 ± 0.012	0.495 ± 0.027

Result analysis: Synthetic datasets

We summarize the CC results of all the methods on four different datasets in Tables 1-3 to illustrate the effectiveness of our proposed method for datasets with different numbers of groups, different number of features, as well as different numbers of samples.

From Tables 1 and 2, we observe that our proposed method outperformed all alternative methods under comparison on different kinds of datasets such as $d \leq n$ in Table 1 and $d \geq n$ in Table 2. For example, our method improved on average by 11.3% and 7.9, respectively, compared with the worst comparison method (*i.e.,* ConRegression) and the best comparison methods (*i.e.,* MKL) on two synthetic datasets. These results indicate that our assumption (*i.e.,* making use of sex specific and common patterns) is reasonable and our proposed method is robust.

Table 3 demonstrates that our method also outperformed all alternative methods under comparison on 2 two-task datasets with the same number of groups as the PNC data. Firstly, feature selection methods (such as Lasso, ConLass, MixLasso, MKL, and our proposed method) outperformed regression methods using all the features (such as Regression and Con-Regression), indicating that the feature selection methods were robust to noisy features in the synthetic datasets. Secondly, different groups had different regression performance, highlighting differences among different groups. Thirdly, concatenation methods (*i.e.,* ConRegression and ConLasso) did not always outperform the corresponding single-task regression methods since heterogeneous data may degrade the prediction performance. These observations indicate that the prediction performance could be improved only if all prediction tasks are modeled effectively, as demonstrated by the results of MKL and our method in Tables 2-3.

Result analysis: PNC data

Sex differences detected by the single prediction model We summarize the prediction results of Regression and Lasso in Table 4, highlighting the developmental sex differences in RSFC measures. Moreover, the age prediction for males outperformed the age prediction for females in terms of the values of coefficient correlation.

In our experiments, we ran our algorithm 200 times (via repeating 10-fold cross validation 20 times) and in each run we kept the features with nonzero coefficients for Lasso. We calculated the frequency of each feature appearing in all these 200 experiments and identify top selected features, *i.e.,* top 50, for each group. We visualize the common patterns among two groups and sex specific RSFC patterns, comprising the selected RSFC features in Figure 1. As shown in Figure 1, Default, Visual, and Frontoparietal networks were sex common networks, while the prediction of age for females selected Visual networks with a higher frequency and the prediction of age for males selected Frontoparietal more frequently. Moreover, Limbic was the least frequently selected network for both the males and females. All the above results are largely consistent with findings in existing studies (Yeo et al. 2011; Satterthwaite et al. 2013).

Sex differences detected by the prediction models in the multi-task learning framework According to Table 4, our method achieved the best prediction performance, followed by MKL, MixLasso, ConLasso, ConRegression, Lasso, and Regression. For example, our method improved on average by 12.35% and 5.82%, respectively, over Lasso and MKL. We also carried out non-parametric paired-sample tests (at 95% significance level) between the results of our method and the results of each alternative method under comparison, and the results indicated that our proposed method outperformed other methods with statistical significance. These results indicate that multi-task learning (such as MKL and our method) may make better use of sex specific and common patterns for predicting age. Moreover, our method outperformed MKL since our method takes into consideration the importance of the groups and the balance of sex specific and common patterns.

Our method in Eq. (4) simultaneously generated two prediction models (*i.e.,* $\mathbf{w}_1$ and $\mathbf{w}_2$, one for each group). Moreover, these two coefficient vectors were sparse. Accordingly, we calculated the frequency of the selected features, and then visualize the corresponding brain networks in Figure 2. From Figure 2, we observe that our method identified RSFC patterns similar to the results in single group. For example, our method also selected RSFC measures among Default, Visual, and Frontoparietal networks as the most important features for the age prediction. In contrast to the previous results, the RSFC measures identified by our method were selected with a higher frequency, indicating that our method had higher reliability than the single group method.

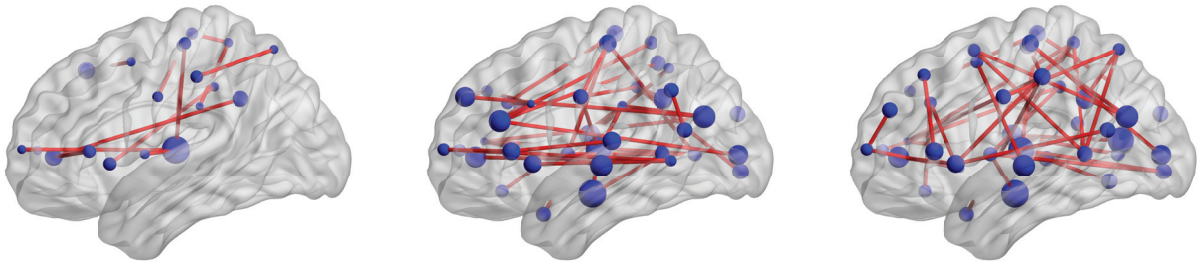
In summary, our method achieved the best prediction performance as well as identified RSFC patterns underlying the developmental sex difference with improved reliability.

Discussion and conclusions

Parameter sensitivity We have shown the variations of CC of our method at different settings of the parameter γ in Figure 3. We observe that the best performance of our method improved on average by 10 compared with the worst case at different datasets, and our method achieved its best results when the sparsity ratio was about $30\% \sim 60\%$. These observations indicate that our method is sensitive to the parameter setting. In our method, we use a square root loss function and an $\ell_{2,1}$-norm regularization to reduce two more parameters to make our algorithm fast. Our method circumvents the parameter tuning problem.

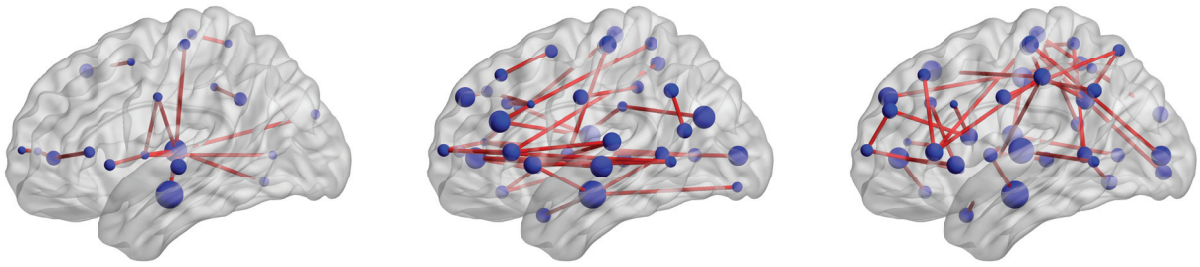
Convergence analysis In Section *Optimization*, we have theoretically proved the convergence of the proposed algorithm for solving Eq. (4). Figure 4 experimentally demonstrates the convergence of the proposed Algorithm 1, showing the objective function values of Eq. (4) on the iteration steps until Algorithm 1 converges. These experimental results indicate that the proposed Algorithm 1 can effectively tackle Eq. (4) with a fast convergence within tens of iteration steps.

In conclusion, we proposed a novel parameter-free centralized multi-task learning method to automatically learn the importance of different tasks and the balance of the distinct and common information associated with the tasks. The experimental results based on both synthetic and real rsfMRI data have demonstrated that our method outperformed the alternative methods under comparison in term of the prediction performance. Furthermore, our method could also



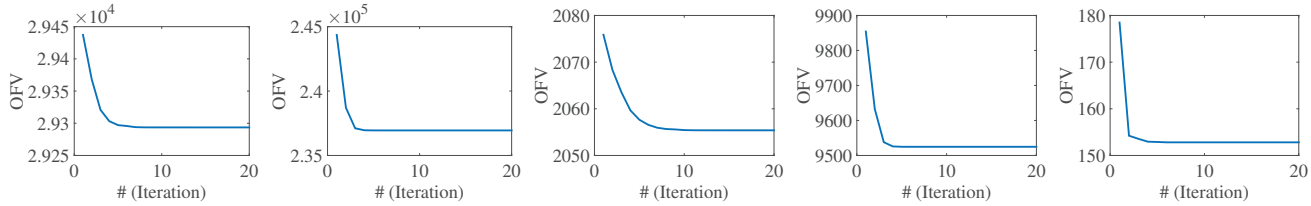
(a) Common networks (b) Males > Females (c) Females > Males

Figure 1: Visualization of common networks and individual specific networks of two single groups selected by Lasso.



(a) Common networks (b) Males > Females (c) Females > Males

Figure 2: Visualization of common networks and individual specific networks of two-task group selected by our method.



(a) Data 1 (b) Data 2 (c) Data 3 (d) Data 4 (e) PNC

Figure 3: The convergence of our proposed Algorithm 1, where the abbreviation "OFV" means "objective function value".

identify informative RSFC patterns that are predictive for the brain development.

References

Alarcón, G.; Cservenka, A.; Rudolph, M. D.; Fair, D. A.; and Nagel, B. J. 2015. Developmental sex differences in resting state functional connectivity of amygdala sub-regions. *NeuroImage* 115:235–244.

Argyriou, A.; Evgeniou, T.; and Pontil, M. 2007. Multi-task feature learning. In *NIPS*, 41–48.

Blakemore, S.-J.; Burnett, S.; and Dahl, R. E. 2010. The role of puberty in the developing adolescent brain. *Human brain mapping* 31(6):926–933.

Chang, X.; Nie, F.; Yang, Y.; and Huang, H. 2014. A convex formulation for semi-supervised multi-label feature selection. In *AAAI*, 1171–1177.

Daubechies, I.; DeVore, R.; Fornasier, M.; and Güntürk, C. S. 2010. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics* 63(1):1–38.

Di Martino, A.; Fair, D. A.; Kelly, C.; Satterthwaite, T. D.; Castellanos, F. X.; Thomason, M. E.; Craddock, R. C.; Luna, B.; Leventhal, B. L.; Zuo, X.-N.; et al. 2014. Unraveling the miswired connectome: a developmental perspective. *Neuron* 83(6):1335–1353.

Fan, Y., and Davatzikos, C. 2017. Pattern recognition of functional brain networks. In *ICASSP*, 6309–6313.

Fox, M. D., and Raichle, M. E. 2007. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature reviews. Neuroscience* 8(9):700.

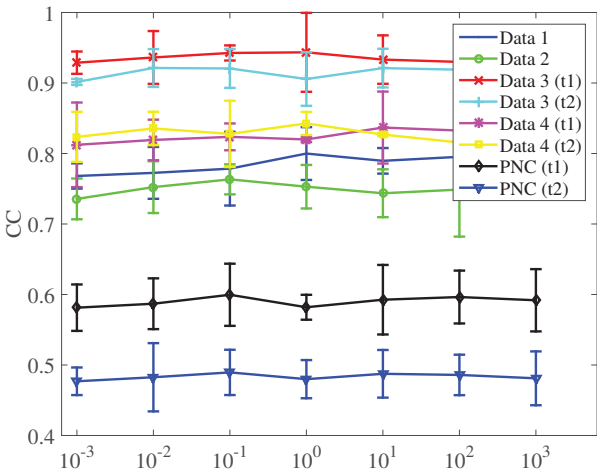


Figure 4: Parameter sensitivity on γ of our method at different datasets. Note that PNC (t1) indicates the results of PNC task 1.

Friedman, J.; Hastie, T.; and Tibshirani, R. 2001. *The elements of statistical learning*, volume 1.

Gur, R. E., and Gur, R. C. 2016. Sex differences in brain and behavior in adolescence: Findings from the philadelphia neurodevelopmental cohort. *Neuroscience & Biobehavioral Reviews* 70:159–170.

Hardee, J. E.; Cope, L. M.; Munier, E. C.; Welsh, R. C.; Zucker, R. A.; and Heitzeg, M. M. 2017. Sex differences in the development of emotion circuitry in adolescents at risk for substance abuse: a longitudinal fmri study. *Social Cognitive and Affective Neuroscience* nsx021.

Hu, R.; Zhu, X.; Cheng, D.; He, W.; Yan, Y.; Song, J.; and Zhang, S. 2017. Graph self-representation method for unsupervised feature selection. *Neurocomputing* 220:130–137.

Ingalhalikar, M.; Smith, A.; Parker, D.; Satterthwaite, T. D.; Elliott, M. A.; Ruparel, K.; Hakonarson, H.; Gur, R. E.; Gur, R. C.; and Verma, R. 2014. Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences* 111(2):823–828.

Kochhann, R.; Gonçalves, H. A.; Pureza, J. d. R.; Viapiana, V. F.; Fonseca, F. d. P.; Salles, J. F.; and Fonseca, R. P. 2017. Variability in neurocognitive performance: Age, gender, and school-related differences in children and from ages 6 to 12. *Applied Neuropsychology: Child* 1–9.

Li, H.; Satterthwaite, T. D.; and Fan, Y. 2017. Large-scale sparse functional networks from resting state fmri. *NeuroImage* 156:1–13.

Müller-Oehring, E. M.; Kwon, D.; Nagel, B. J.; Sullivan, E. V.; Chu, W.; Rohlfing, T.; Prouty, D.; Nichols, B. N.; Poline, J.-B.; Tapert, S. F.; et al. 2017. Influences of age, sex, and moderate alcohol drinking on the intrinsic functional architecture of adolescent brains. *Cerebral Cortex* 1–15.

Peng, H., and Fan, Y. 2016. Direct sparsity optimization based feature selection for multi-class classification. In *IJCAI*, 1918–1924.

Peng, H., and Fan, Y. 2017a. Feature selection by optimizing a lower bound of conditional mutual information. *Information Sciences* 418:652–667.

Peng, H., and Fan, Y. 2017b. A general framework for sparsity regularized feature selection via iteratively reweighted least square minimization. In *AAAI*, 2471–2477.

Qian, B.; Wang, X.; Cao, N.; Li, H.; and Jiang, Y.-G. 2015a. A relative similarity based method for interactive patient risk prediction. *Data Mining and Knowledge Discovery* 29(4):1070–1093.

Qian, B.; Wang, X.; Ye, J.; and Davidson, I. 2015b. A reconstruction error based framework for multi-label and multi-view learning. *IEEE Transactions on Knowledge and Data Engineering* 27(3):594–607.

Satterthwaite, T. D.; Elliott, M. A.; Gerraty, R. T.; Ruparel, K.; Loughead, J.; Calkins, M. E.; Eickhoff, S. B.; Hakonarson, H.; Gur, R. C.; Gur, R. E.; et al. 2013. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage* 64:240–256.

Satterthwaite, T. D.; Elliott, M. A.; Ruparel, K.; Loughead, J.; Prabhakaran, K.; Calkins, M. E.; Hopson, R.; Jackson, C.; Keefe, J.; Riley, M.; et al. 2014. Neuroimaging of the philadelphia neurodevelopmental cohort. *Neuroimage* 86:544–553.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 267–288.

Yang, Y.; Ma, Z.; Nie, F.; Chang, X.; and Hauptmann, A. G. 2015. Multi-class active learning by uncertainty sampling with diversity maximization. *International Journal of Computer Vision* 113(2):113–127.

Yeo, B. T.; Krienen, F. M.; Sepulcre, J.; Sabuncu, M. R.; Lashkari, D.; Hollinshead, M.; Roffman, J. L.; Smoller, J. W.; Zöllei, L.; Polimeni, J. R.; et al. 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology* 106(3):1125–1165.

Zhang, S.; Li, X.; Zong, M.; Zhu, X.; and Wang, R. 2017. Efficient knn classification with different numbers of nearest neighbors. *IEEE Transactions on Neural Networks and Learning Systems, DOI: 10.1109/TNNLS.2017.2673241*.

Zhou, J.; Chen, J.; and Ye, J. 2011. Malsar: Multi-task learning via structural regularization.

Zhu, X.; Suk, H.-I.; Lee, S.-W.; and Shen, D. 2016a. Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification. *IEEE Transactions on Biomedical Engineering* 63(3):607–618.

Zhu, Y.; Zhu, X.; Kim, M.; Shen, D.; and Wu, G. 2016b. Early diagnosis of alzheimers disease by joint feature selection and classification on temporally structured support vector machine. In *MICCAI*, 264–272.

Zhu, X.; Suk, H.-I.; Huang, H.; and Shen, D. 2017a. Low-rank graph-regularized structured sparse regression for identifying genetic biomarkers. *IEEE Transactions on Big Data, DOI: 10.1109/TBDATA.2017.2735991*.

Zhu, X.; Suk, H.-I.; Wang, L.; Lee, S.-W.; Shen, D.; Initiative, A. D. N.; et al. 2017b. A novel relational regularization feature selection method for joint regression and classification in ad diagnosis. *Medical image analysis* 38:205–214.

Zhu, Y.; Zhu, X.; Kim, M.; Yan, J.; and Wu, G. 2017c. A tensor statistical model for quantifying dynamic functional connectivity. In *IPMI*, 398–410.

Zhu, X.; Suk, H.-I.; and Shen, D. 2014. Multi-modality canonical feature selection for alzheimers disease diagnosis. In *MICCAI*, 162–169.